SOURCE AND ACCURACY STATEMENT FOR THE SURVEY OF PROGRAM DYNAMICS 2001 CROSS-SECTIONAL FILE

DATA COLLECTION AND ESTIMATION

Source of Data

The Survey of Program Dynamics (SPD) universe is the noninstitutionalized resident population living in the United States between 1992 and 2002. This population includes people (including children) living in group quarters, such as dormitories, rooming houses, and religious group dwellings. Crew members of merchant vessels, Armed Forces personnel living in military barracks, and institutionalized people, such as correctional facility inmates and nursing home residents, were not eligible to be in the survey. In addition, United States citizens residing abroad were not eligible to be in the survey. Foreign visitors who work or attend school in this country and their families were eligible; all others were not eligible to be in the survey. With the exceptions noted above, people who were at least 15 years of age at the time of the interview were eligible to be asked questions regarding income and job experience.

The goal of the SPD program is to provide policy makers with data to assess the effects of the welfare reforms enacted by the U.S. Congress in 1996 and how these reforms interact with each other, and with employment, income, and family circumstances. The SPD program spans from the pre-reform through the post-reform period, 1992-2002. In order to obtain information about past economic history, employment, income, and program participation prior to the welfare reforms, two retired SIPP panels 1992 and 1993 were chosen as the SPD samples fielded between 1997 and 2002. Each of these SPD samples consisted of sample households selected from SIPP Panels 1992 and 1993 as described later (up to SPD 2001 in this document, i.e., the source and accuracy statement for the SPD 2001 cross-sectional file).

The SPD samples were fielded once a year either between the months of April and June or between the months of May and July from 1997 through 2002. The first SPD sample was fielded in 1997 and is referred to as the SPD Bridge. Each SPD sample collected annual socioeconomic and health insurance data from the previous year from the sample households but collected demographic data for the current year similar to the March supplement of the Current Population Survey (CPS). Therefore, the March CPS instrument (questionnaire) was used directly for collecting data from the SPD Bridge sample. For the data collection of other SPD samples (SPD 1998 to SPD 2002), the March CPS instrument was used with necessary modifications, mainly to accommodate additional topics included in the data collection. The combined data from SIPP Panels 1992 and 1993, SPD Bridge sample, and SPD 1998 to 2002 samples should yield the necessary pre-welfare-reform and post-welfare-reform information for assessing the effects of welfare reform on the socioeconomic well-being of the U.S. non-institutionalized population. The full potential of the SPD data is generally achieved when using the SPD cross-sectional and longitudinal files. The current source and accuracy statement is provided for the SPD 2001 cross-sectional file.

Background of SIPP 1992 and 1993 Panels and SPD Bridge Survey

The 1992 and 1993 SIPP panel samples were located in 284 Primary Sampling Units (PSUs), each consisting of a county or a group of contiguous counties. Within these PSUs, expected clusters of two or four living quarters (LQs) were systematically selected from lists of addresses prepared for the 1980 decennial census to form the bulk of the sample. To account for LQs built within each of the sample areas after the 1980 census, a sample was drawn of permits issued for construction of residential LQs up until shortly before the beginning of the panel. In jurisdictions that do not issue building permits, small land areas were sampled and the LQs within were listed by field personnel and then sub-sampled. In addition, sample LQs were selected from supplemental frames that included LQs identified as missed in the 1980 census and group quarters (GQs).

At the time of the initial visit of the SIPP panels, the occupants of about 19,600 living quarters were interviewed for the 1992 panel and 19,900 were interviewed for the 1993 panel. This accounts for approximately 72% (1992) and 73% (1993) of the LQs originally designated for the SIPP samples. Approximately 21% (1992) and 20% (1993) of the designated LQs were found to be vacant, demolished, converted to nonresidential use, or otherwise ineligible for the survey. The remainder, approximately 2000 LQs, were not interviewed because the occupants refused to be interviewed, could not be found at home, were temporarily absent, or otherwise unavailable. Thus, occupants of about 91% of all eligible LQs participated in the first interview of the 1992 and 1993 SIPP panels.

For the remaining nine interviews, only original sample people (those in Wave 1 sample households and interviewed in Wave 1) and people living with them were eligible to be interviewed. With certain restrictions, original sample people were to be followed even if they moved to a new address. When original sample people moved without leaving a forwarding address or moved to extremely remote parts of the country with no telephone number available for contact, additional non-interviews resulted.

The 1992 ten-wave cross-sectional and longitudinal files from SIPP Panel 1992 consist of data collected from February 1992 to April 1995. Data for up to 39 reference months are available for people on these files. The 1993 nine-wave cross-sectional and longitudinal files from SIPP Panel 1993 consist of data collected from February 1993 to January 1996. Data for up to 36 reference months are available for people on these files.

Table 1 indicates the interview and reference months for the collection of data from SIPP Panel 1992 (ten waves), SIPP Panel 1993 (nine waves), and SPD1997 to 2002. For the SIPP, a person was classified as interviewed or non-interviewed based on the following definitions (note: a person may be classified differently for calculating different weights). Interviewed sample people (including children) were defined to be: "those for whom self or proxy responses were obtained for the month or months (e.g., a longitudinal period) under consideration". The month or months under consideration for which people were deceased or residing in an ineligible address were identified on both the cross-sectional and longitudinal files. Non-interviewed people were defined to be those for whom neither self nor proxy responses were obtained for the month or months under consideration. However, for the weighting purpose, non-interviewed sample people whose responses were obtained by imputation for the month or months under consideration were reclassified as interviewed.

It is estimated that roughly 21,600 households (56,300 people) for SIPP Panel 1992 and 21,800 households (57,200 people) for SIPP Panel 1993 were initially designated in the sample for the SIPP interview. Approximately 19,600 households (51,100 people) for SIPP Panel 1992 and 19,900 households (51,900 people) for SIPP Panel 1993 were interviewed in wave 1. The balance, residing in approximately 4,000 (SIPP Panels 1992 and 1993 combined) living quarters not interviewed at wave 1, remained anonymous and became the initial source of the household (and person) non-response in the weighting procedures. For wave 2 and beyond, household non-response was due to *Type A non-interview - direct or indirect refusal for interview*, *Type B non-interview - entire household institutionalized*, *Type C non-interview - entire household deceased or moved out of the SPD universe* (e.g., out of the country), and *Type D non-interview - unlocated mover's household*. By the end of the panel, with sample growth due to household spawning outpacing the household non-response, roughly 23,600 households for SIPP Panel 1992 (waves 9 and 10) and 23,700 households for SIPP Panel 1993 (wave 9) were eligible for interview; and approximately 17,800 households for SIPP Panel 1992 and 17,600 households for SIPP Panel 1993 were interviewed. In an interviewed household, some household members refused to be interviewed which is referred to as Type Z non-response. For SIPP Panels 1992 and 1993, Type Z non-response rates varied from 1.3% to 8.0% among wave 2 and later waves, and the rates increased as the panels aged. In addition, respondents did not respond to some of the questions; therefore, item non-response rates, especially for sensitive income and money related items, are higher than the person non-response rates given above.

We define the SPD Bridge sample cohort as people that were in an interviewed household in the last waves of SIPP Panels 1992 and 1993 plus the people in the households spawned from the interviewed households in the last waves of SIPP Panel 1992 and 1993. This amounted to approximately 38,300 designated sample households for the SPD Bridge sample among which 30,125 households were interviewed. Among the non-interviewed households, 4,335 were Type A, 302 Type B, 1,181 Type C, and 2,402 Type D non-interviewed households.

1998 SPD Sub-sampling for SPD 1998 Survey

Due to budget constraints, the SPD 1998 Survey did not visit all the households eligible for interview (approximately 39,200 interviewed, Type A, Type B, and Type D non-interviewed households in the SPD Bridge plus their spawned households). The budget only allowed for SPD to visit 21,000 households. Thus, a sample cut was carried out in the following manner. Roughly 19,100 were sampled in this operation, since we needed to account for an expected 12.5 percent non-response and a growth of 10 percent of the total sample size due to household spawning. Only the interviewed and Type B non-interviewed households in the SPD Bridge sample were eligible for selection for the SPD 1998 sample. In the sub-sampling (sample selection), these sample households were demographically divided into six strata as shown in the table below. The stratification was performed using the household information collected from the SPD Bridge. In each stratum, the households were randomly selected with the sampling rate as provided in the table below. As indicated among sampling rates in this table, the low income sample households were generally not subjected to the sample cut at all.

As a result of the sample cut, the number of the households selected for the SPD 1998 sample plus their spawned households was approximately 19,700 among which 16,395 households were interviewed. Among the non-interviewed households, 2,211 were Type A, 201 Type B, 212 Type C, and 673 Type D non-interviewed households.

Stratum	Description	Sampling Rate	Designated Number	Projected Interviews
1	Households where the primary family or the primary individual has a total family income below 150% of the poverty threshold	1-in-1	6,182	5,950
2	Households where the primary family or the primary individual has a total family income between 150% and 200% of the poverty threshold and there are children under 18	1-in-1	1,075	1,035
3	Households where the primary family or the primary individual has a total family income above 200% of the poverty threshold and there are children under 18	1-in-1.11	6,623	6,375
4	Households where the primary family or the primary individual has a total family income between 150% and 200% of the poverty threshold and there are no children under 18	1-in-1.22	1,461	1,406
5	Households in the balance	1-in-3.70	3,707	3,568
6	Households entirely institutionalized (Outcome code = 228)	1-in-3.70	81	DK
Total			19,129	18,334

SPD 1999 Survey

The sample for the SPD 1999 survey is simply a direct continuation of the SPD 1998 sample. Namely, there is no additional sample cut nor bringing back of any additional sample households from either the SPD Bridge non-respondents or the SIPP Panel 1992/1993 non-respondents. In this survey, there are approximately 20,000 designated sample households (including the spawned households) among which 16,395 households were interviewed. Among the non-interviewed households, 2,174 were Type A, 232 Type B, 212 Type C, and 718 Type D non-interviewed households.

SPD 2000 Survey

The sample for the SPD 2000 survey consists of two parts. Part 1 of the SPD 2000 sample is simply a direct continuation of the SPD 1999 sample which are hereinafter referred to as *the SPD core part* or *the SPD basic part*. Part 2 is composed of a sub-sample of the SPD Bridge Type A and Type D non-interviewed households which are hereinafter referred to as *the SPD Bridge non-interview part*.

The SPD Bridge non-interview part of the sample was selected by classifying the Type A and Type D SPD Bridge non-interviewed households into six strata similar to those prescribed for the 1998 SPD sample cut as shown in the table below. The non-interviewed households in each stratum were then randomly selected at the sampling rate specified in the table below (which is the same sampling rate for the 1998 SPD sample cut).

Stratum	Description[1]	Number of Households	Percent	Sampling Rate	Sub-sample Size
1	Households where the primary family or the primary individual has a total family income below 150% of the poverty threshold	1,444	22.5	1-in-1	1,444
2	Households where the primary family or the primary individual has a total family income between 150% and 200% of the poverty threshold and there are children under 18	156	2.4	1-in-1	156
3	Households where the primary family or the primary individual has a total family income above 200% of the poverty threshold and there are children under 18	736	11.5	1-in-1.11	663
4	Households where the primary family or the primary individual has a total family income between 150% and 200% of the poverty threshold and there are no children under 18	476	7.4	1-in-1.22	390
5	Households in the balance	3,605	56.2	1-in-3.70	974
6	Households institutionalized (outcome code = 228)	0	0.0	1-in-3.70	0
Total		6,417	100.0	N.A.	3,627

[1] The characteristics of these households were determined based on Reference Month 4 of Wave 1 of either the SIPP Panel 1992 or 1993, depending on which panel they were originated from.

In the SPD core part of the sample, there are approximately 20,400 designated sample households (including the spawned households) among which 16,845 households were interviewed. Among the non-interviewed households in the SPD core part, 2,172 were Type A, 283 Type B, 243 Type C, and 873 Type D non-interviewed households. In the SPD Bridge non-interview part of the sample, there are approximately 3,700 designated sample households (including the spawned households) among which 1,871 households were interviewed. Among the non-interviewed households in the SPD Bridge non-interview part, 806 were Type A, 32 Type B, 98 Type C, and 907 Type D non-interviewed households. Thus, for the whole SPD 2000, there are approximately 24,100 designated sample households (including the spawned households) among which 18,716 households were interviewed. Among the non-interviewed households in the whole SPD 2000 sample, 2,978 were Type A, 315 Type B, 341 Type C, and 1,780 Type D non-interviewed households.

SPD 2001 Survey

The sample for the SPD 2001 survey consists of three parts. Part 1 of the SPD 2000 sample is simply a direct continuation of the SPD core part of the SPD 2000 sample. Part 2 is also a direct continuation of the SPD Bridge non-interview part of the SPD 2000 sample. Part 3 is composed of SIPP Panel 1992 and 1993 Type A and Type D non-interviewed households between wave 2 and the last wave. Part 3 of the SPD 2001 sample will be referred to as *the SPD 92/93 non-interviewed part*.

The SPD 92/93 non-interview part of the sample was selected by classifying the SIPP Panel 1992 and 1993 Type A and Type D SPD non-interviewed households (between wave 2 and the last wave) into six strata (similar to those prescribed for the 1998 SPD sample cut) as shown in the table below. The non-interviewed households in each stratum were then randomly selected at the sampling rate specified in the table below.

In the SPD core part of the sample, there are approximately 20,900 designated sample households (including the spawned households) among which 16,964 households were interviewed. Among the non-interviewed households in the SPD core part, 2,351 were Type A, 314 Type B, 257 Type C, and 1,023 Type D non-interviewed households. In the SPD Bridge non-interview part of the sample, there are approximately 3,800 designated sample households (including the spawned households) among which 2,272 households were interviewed. Among the non-interviewed households in the SPD Bridge non-interview part, 816 were Type A, 42 Type B, 50 Type C, and 655 Type D non-interviewed households. In the SPD 92/93 non-interview part, there are approximately 6,400 designated sample household (including the spawned households) among which 3,104 households were interviewed. Among the non-interviewed households in the SPD 92/93 non-interview part, 1,056 were Type A, 85 Type B, 240 Type C, and 1907 Type D non-interviewed households. Thus, for the whole SPD 2001, there are approximately 31,100 designated sample households (including the spawned households) among which 22,340 households were interviewed. Among the non-interviewed households in the whole SPD 2001 sample, 4,223 were Type A, 441 Type B, 547 Type C, and 3,385 Type D non-interviewed households.

Stratum	Description[1]	Number of Households		Percent		Sampling Rate		Sub-sample Size	
		1992	1993	1992	1993	1992	1993	1992	1993
1	Households where the primary family or the primary individual has a total family income below 150% of the poverty threshold	1317	1463	31.1	32.2	1 in 1	1 in 1	1317	1463
2	Households where the primary family or the primary individual has a total family income between 150% and 200% of the poverty threshold and there are children under 18	182	254	4.3	5.6	1 in 1	1 in 1	182	254
3	Households where the primary family or the primary individual has a total family income above 200% of the poverty threshold and there are children under 18	723	807	17.0	17.7	1 in 1	1in 1	723	807
4	Households where the primary family or the primary individual has a total family income between 150% and 200% of the poverty threshold and there are no children under 18	270	293	6.4	6.4	1 in 1	1 in 1	270	293
5	Households in the balance	1746	1733	41.2	38.1	1 in 3.44	1 in 9.47	508	183
Total by Panel		4238	4550	100.0	100.0	N.A.	N.A.	3000	3000
Grand Total (1992 & 1993 Panels)		8788		N.A.		N.A.		6000	

[1] The characteristics of these households were determined based on Reference Month 4 of Wave 1 of either the SIPP Panel 1992 or 1993, depending on which panel they originally belonged.

ESTIMATION

For estimation of the characteristics of a cohort of people in the SPD 2001 universe, use the final cross-sectional weights (SPDCSW01) of the sample persons in the cohort provided on the SPD 2001 cross-sectional file. The procedure for calculating the final cross-sectional weights for the sample persons was described earlier in the Weighting section. All sample persons classified as interviewed persons in the SPD 2001 were assigned a positive SPD 2001 final cross-sectional weight (SPDCSW01 > 0), while all those classified otherwise were assigned a zero SPD 2001 final cross-sectional weight (SPDCSW01 = 0). A sample person is classified as an interviewed person in the SPD 2001 if he/she is a self, proxy, or imputed respondent in the SPD 2001. As mentioned earlier, on the SPD 2001 cross-sectional file, the annual socioeconomic and health insurance data of the sample people was from the previous year (2000) but the demographic data of the sample people was from the current year (2001). Namely, the SPD 2001 cross-sectional file provides the micro-data for estimations of the socioeconomic characteristics in 2000 and demographic characteristics in 2001 for the noninstitutionalized people who are resident in the United States in 2001.

For estimation of the characteristics of a cohort of households in the SPD 2001 universe, use the final cross-sectional weights (SPDCSW01) of the reference persons of the sample households in the cohort on the SPD 2001 cross-sectional file. As with the person level data, on the SPD 2001 cross-sectional file, the annual socioeconomic and health insurance data of the sample households were from the previous year (2000) but the demographic data of the sample households were from the current year (2001). Namely, the SPD 2001 cross-sectional file provides the micro-data for estimations of the socioeconomic characteristics in 2000 and demographic characteristics in 2001 of the households of the noninstitutionalized people who are resident in the United States in 2001.

For estimation of the characteristics of a cohort of families (primary family and/or subfamily in a household) in the SPD 2001 universe, use the final cross-sectional weights (SPDCSW01) of the reference persons of the sample families (primary families and/or subfamilies in the sample households) in the cohort on the SPD 2001 cross-sectional file. As with the person level data, on the SPD 2001 cross-sectional file, the annual socioeconomic and health insurance data of the sample families were from the previous year (2000) but the demographic data of the sample families were from the current year (2001). Namely, the SPD 2001 cross-sectional file provides the micro-data for estimations of the socioeconomic characteristics in 2000 and demographic characteristics in 2001 of the families of the noninstitutionalized people who are resident in the United States in 2001.

ACCURACY OF ESTIMATES

SPD estimates are based on a sample; they may differ somewhat from the figures that would have been obtained if a complete census had been taken using the same questionnaire, instructions, and enumerators. There are two types of errors possible in an estimate based on a sample survey: non-sampling and sampling. We are able to provide estimates of the magnitude of SPD sampling error, but this is not true of non-sampling error. The next sections describe sources of SPD non-sampling error, followed by a discussion of sampling error, its estimation, and its use in data analysis.

Note that estimates from this sample for individual states are subject to very high sampling errors and are not recommended because all SPD samples were derived from SIPP Panel 1992 and 1993 samples which were designed to be nationally representative but not state representative. The state codes on the file are primarily of use for linking respondent characteristics with appropriate contextual variables (e.g., state specific welfare criteria) and for tabulating data by user-defined groupings of states.

Non-sampling Errors

Non-sampling errors can be attributed to many sources: for example, inability to obtain information about all cases in the sample, difficulties in precisely stating some definitions, differences in the interpretation of questions, inability or unwillingness on the part of the respondents to provide correct information, inability to recall information, and errors of collection such as in recording or coding the data, processing the data, estimating values for missing data, biases resulting from the differing recall periods caused by the rotation pattern used, and under-coverage. Quality control and edit procedures were used to reduce errors made by respondents, coders and interviewers.

Under-coverage in SPD results from missed living quarters and missed people within sample households. It is known that under-coverage varies with age, race, and gender. Generally, under-coverage is larger for males than for females and larger for Blacks than for non-Blacks. Ratio estimation to independent age-race-gender population controls (benchmark estimates) partially corrects for the bias due to survey under-coverage. However, biases exist in the estimates to the extent that people in missed households or missed people in interviewed households have characteristics different from those of interviewed people in the same age-race-gender group. In addition, the independent population controls used have not been adjusted for under-coverage in the decennial census. The Census Bureau has used complex techniques to adjust the weights for non-response. For an explanation of the techniques used, see the “Non-response Adjustment Methods for Demographic Surveys at the U.S. Bureau of the Census,” November 1988, Working Paper 8823, by R. Singh and R. Petroni. An example of successfully avoiding bias can be found in "Current Non-response Research for the Survey of Income and Program Participation" (paper by Petroni, presented at the Second International Workshop on Household Survey Non-response, October 1991). The procedure for calculating the person cross-sectional weights on the SPD 2001 cross-sectional file was derived based on such complex techniques.

The SIPP 2001 cross-sectional sample may not perform as well as the CPS March 2001 sample and SIPP Panel 2001 sample in representing the noninstitutionalized national population in 2001. This is principally attributable to the fact that, as the SPD sample aged more, it became less representative of the current population (such as the 2001 national population). In addition, the high sample loss (attrition) rate (roughly 34 percent) in the whole SPD 2001 sample may reduce the degree of the effectiveness of the non-interview adjustment process to fully compensate for the differential attrition. Note that the sample loss rate for SPD 2001 has two components: 27 percent sample loss inherited from the SIPP Panels 1992 and 1993, and additional net sample loss of 7 percent occurred from the SPD 1997 (Bridge) to 2001 offset by the bringing-back of the SPD Bridge non-interview and SIPP 92/93 non-interview sample households as discussed earlier.

Comparability with Other Estimates

Caution should be exercised when comparing data from this file with data from SIPP publications or with data from other surveys, such as the Current Population Survey (CPS). The comparability problems are caused by such sources as the seasonal patterns for many characteristics, different non-sampling errors, and different concepts and procedures. Refer to the *SIPP Quality Profile* for known differences with data from other sources and further discussion.

Sampling Variability

Standard errors indicate the magnitude of the sampling error. They also partially measure the effect of some non-sampling errors in response and enumeration, but do not measure any systematic biases in the data. The standard errors for the most part measure the variations that occurred by chance because a sample rather than the entire population was surveyed.

USES AND COMPUTATION OF STANDARD ERRORS

Confidence Intervals

The sample estimate and its standard error enable one to construct confidence intervals (ranges that would include the average result of all possible samples with a known probability). For example, if all possible samples were selected, each of these being surveyed under essentially the same conditions and using the same sample design, and if an estimate and its standard error were calculated from each sample, then:

1. Approximately 90 percent of the intervals from 1.645 standard errors below the estimate to 1.645 standard errors above the estimate would include the average result of all possible samples.

2. Approximately 95 percent of the intervals from 1.960 standard errors below the estimate to 1.960 standard errors above the estimate would include the average result of all possible samples.

The average estimate derived from all possible samples is or is not contained in any particular computed interval. However, for a particular sample, one can say with a specified confidence that the average estimate derived from all possible samples is included in the confidence interval.

Hypothesis Testing

Standard errors may also be used for hypothesis testing, a procedure for distinguishing between population characteristics using sample estimates. The most common types of hypotheses tested are the population characteristics being identical versus being different. Tests may be performed at various levels of significance, where a level of significance is the probability of concluding that the characteristics are different when, in fact, they are identical.

To perform the most common test, compute the difference X_A - X_B, where X_A and X_B are sample estimates of the characteristics of interest. A later section explains how to derive an estimate of the standard error of the difference X_A - X_B. Let that standard error be s_{DIFF}. If X_A - X_B is between

-1.645 times s_{DIFF} and +1.645 times s_{DIFF}, no conclusion about the characteristics is justified at the 10 percent significance level. If, on the other hand, X_A - X_B is smaller than -1.645 times s_{DIFF} or larger than +1.645 times s_{DIFF}, the observed difference is significant at the 10 percent level. In this event, it is commonly accepted practice to say that the characteristics are different. We recommend that users report only those differences that are significant at the 10 percent level or better. Of course, sometimes this conclusion will be wrong. When the characteristics are, in fact, the same, there is a 10 percent chance of concluding that they are different.

Note that as more tests are performed, more erroneous significant differences will occur. For example, at the 10 percent significance level, if 100 independent hypothesis tests are performed in which there are no real differences, it is likely that about 10 erroneous differences will occur. Therefore, the significance of any single test should be interpreted cautiously. However, we can generally increase the power of the test by using a multiple significance test (multiple comparison procedure) in lieu of a series of independent pair tests. For a number of simultaneous tests of five or less, we recommend Bonferroni's procedure which generally provides a simple and acceptable answer to multiple tests. It is however highly conservative and not recommended if the multiple tests are higher than five because many other multiple comparison procedures available (such as least significant difference test, Newman-Keuls test, Scheffe's test, etc.) are generally preferable.

Caution Concerning Small Estimates and Small Differences

Because of the large standard errors involved, there is little chance that estimates will reveal useful information when computed on a base smaller than 200,000. Furthermore, non-sampling error in one or more of the smaller number of sample units on which an estimate is based can be relatively large (unacceptably large) in the estimate (because the non-sampling error will be generally self-cancelled in an estimate if the number of sample units used for the estimate is large). In addition, care must be taken in the interpretation of small differences since even a small amount of non-sampling error can cause a borderline difference to appear significant or not, thus distorting a seemingly valid hypothesis test.

Standard Error Parameters

Most SPD 2001 estimates have greater standard errors than those obtained through a simple random sample because clusters of living quarters are sampled for the SIPP, SPD Bridge, and SPD 1998 to SPD 2001. To derive standard errors that would be applicable to a wide variety of estimates and could be prepared at a moderate cost, a number of approximations were required. Estimates with similar standard error behavior were grouped together and two parameters (denoted *a* and *b*) were developed to approximate the standard error behavior of each group of estimates. Because the actual standard error behavior was not identical for all estimates within a group, the standard errors computed from these parameters provide an indication of the order of magnitude of the standard error for any specific estimate. These *a* and *b* parameters vary by characteristic and by demographic subgroup to which the estimate applies. The *a* and *b* parameters are also known as "generalized variance parameters."

For the SPD 2001 cross-sectional file, the *a* and *b* parameters for various groups of the populations and households are provided in Table 3.

Standard Errors of Estimated Numbers

The approximate standard error s_x of an estimated number x of people, families and so forth, can be obtained by using Formula 1 provided below.

$$s_x = \sqrt{ax^2 + bx} \qquad (1)$$

Here *a* and *b* are the standard error parameters associated with the particular type of characteristic in a given point or period of time. For the analysis using the data on the SPD 2001 cross-sectional file, the *a* and *b* parameters are provided in Table 3 for estimation of socioeconomic characteristics in 2000 and demographic characteristic in 2001.

An illustration would be to suppose that using the SPD 2001 cross-sectional file data, the estimate of the number of people in 2001 ever receiving Social Security in 2000 is 34,122,000 which was derived using the SPD 2001 final cross-sectional weight (SPDCSW01). The appropriate *a* and *b* parameters to use in calculating a standard error for the estimate are obtained from Table 3. They are a = -0.0000615, b = 12,007. Using Formula (1), the approximate standard error s_x is

$$s_x = \sqrt{(-0.0000615)(34{,}122{,}000)^2 + (12{,}007)(34{,}122{,}000)} = 581{,}462 \quad \textit{people}$$

The 90-percent confidence interval as shown by the data is from 33,165,495 to 35,078,505. Therefore, a conclusion that the average estimate derived from all possible samples lies within a range computed in this way would be correct for roughly 90 percent of all samples. Similarly, the 95-percent confidence interval as shown by the data is from 32,982,335 to 35,261,665 and we could conclude that the average estimate derived from all possible samples lies within this interval with 95% confidence.

Standard Error of a Mean and an Aggregate

A mean $\bar{x}$ is defined here to be the average quantity of some characteristic (other than the number of people, families, or households) per person, family, or household. An aggregate k is defined to be the total quantity of some characteristic summed over all units in a sub-population. For example, a mean could be the average annual income of females age 25 to 34. The standard error $s_{\bar{x}}$ of a mean can be approximated by Formula 2 and the standard error s_k of an aggregate can be approximated by Formula 3. Because of the approximations used in developing Formulas 2 and 3, an estimate of the standard error of the mean or aggregate obtained from these formulas will generally underestimate the true standard error. The formula used to estimate the standard error $s_{\bar{x}}$ of a mean $\bar{x}$ is

$$s_{\bar{x}} = \sqrt{\left(\frac{b}{y}\right) s^2} \qquad (2)$$

where y is the base, s^2 is the estimated population variance of the characteristic, and b is the standard error parameter associated with the type of the characteristic.

By defining an aggregate (k) as the total quantity of an item summed over all the units in a group, an approximate standard error (s_k) of the aggregate k can be expressed as Formula 3 below.

$$s_k = \sqrt{by\, s^2} \qquad (3)$$

Because of the approximations used in developing Formulas 3 and 4, an estimate of the standard error of the mean using Formula 2 and an estimate of standard error of the aggregate using Formula 3 will generally underestimate their corresponding true standard errors.

The population variance s^2 may be estimated by one of two methods: the first method uses data that has been grouped into intervals, the second method uses ungrouped data. The second method is recommended because it is more precise. However, the first method will be easier to implement if grouped data are already being used as part of the analysis. In both methods, let x_i denote the value of the characteristic for the i^{th} person.

To use the first method, the range of values for the characteristic is divided into c intervals, where the lower and upper boundaries of interval j are Z_{j-1} and Z_j, respectively. Each person is placed into one of the c groups such that the value of the characteristic, x_i is between Z_{j-1} and Z_j. The estimated population variance, s^2 is then given by Formula 4 below.

$$s^2 = \sum_{j=1}^{c} p_j m_j^2 - \bar{x}^2 \qquad (4)$$

where p_j is the estimated proportion of people in group j (based on weighted data), and m_j is given by Formula 5a at the top of the next page.

$$m_j = \frac{Z_{j-1} + Z_j}{2}, \quad for \quad j = 1,\ 2,\ \ldots,\ c \tag{5a}$$

The most representative value of the characteristic in group j is assumed to be m_j. If group c is open-ended, that is, no upper interval boundary exists, then an approximate value for m_c is given by Formula 5b below.

$$m_c = \left(\frac{3}{2}\right) Z_{c-1} \tag{5b}$$

The mean $\bar{x}$ can be obtained using Formula 6 below.

$$\bar{x} = \sum_{j=1}^{c} p_j m_j \tag{6}$$

In the second method, the estimated population variance s^2 is given by Formula 7 below.

$$s^2 = \frac{\sum_{i=1}^{n} w_i x_i^2}{\sum_{i=1}^{n} w_i} - \bar{x}^2 \tag{7}$$

where there are n sample people with the characteristic of interest and w_i is the final weight for person i. The mean $\bar{x}$ can be obtained from Formula 8 below.

$$\bar{x} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i} \tag{8}$$

Note that, by definition, y (the size of the base) in Formulas 2 and 3 can be obtained from the equation below.

$$y = \sum_{i=1}^{n} w_i$$

An illustration of Method 1 would be to suppose that, based on the SPD 2001 final cross-sectional weight (SPDCSW01), the 2000 distribution of annual incomes of the people in 2001 is given in Table 2 for people aged 25 to 34 who were employed for all 12 months of 2000. The mean annual cash income from Formula 6 is

$$\bar{x} = \frac{1{,}371}{39{,}923}(2{,}500) + \frac{1{,}651}{39{,}923}(6{,}250) + \ \ldots \ + \frac{1{,}493}{39{,}923}(105{,}000) = \$26{,}749$$

Using Formula 4 and the mean annual cash income of \$26,749 the estimated population variance, s^2 is

$$s^2 = \frac{1{,}371}{39{,}923}(2{,}500)^2 + \frac{1{,}651}{39{,}923}(6{,}250)^2 + \ \ldots \ + \frac{1{,}493}{39{,}923}(105{,}000)^2 - 26{,}749^2 = 468{,}092{,}251$$

The appropriate *b* parameter from Table 3 is 6,556. Now, using Formula 2, the estimated standard error of the mean is

$$s_{\bar{x}} = \sqrt{\frac{6{,}556}{39{,}923{,}000}(468{,}092{,}251)} = \$277$$

Standard Errors of Estimated Percentages

This section refers to the percentages of a group of people, families, or households possessing a particular attribute and to percentages of money or related concepts. The reliability of an estimated percentage, computed using sample data for both numerator and denominator, depends upon both the size of the percentage and the size of the total upon which the percentage is based. Estimated percentages are relatively more reliable than the corresponding estimates of the numerators of the percentages, particularly if the percentages are more than 50 percent. For example, the percent estimate of employed people is more reliable than the estimated number of employed people. When the numerator and denominator of the percentage have different parameters, use the parameter of the numerator. If proportions are presented instead of percentages, note that the standard error of a proportion is equal to the standard error of the corresponding percentage divided by 100.

There are two types of percentages commonly estimated. The first type is the percentage of people sharing a particular characteristic such as the percentage of people owning their own home in 2001 or the percentage of food stamp recipients in 2000 among all the people in 2001. The second type is the percentage of money or some similar concept held by a particular group of people or held in a particular form. Examples are the percentage of wealth held by people with high income and the percentage of annual income received by females.

For the percentage of people, the approximate standard error, $s_{x,p}$, of the estimated percentage, p, can be obtained by Formula 9 below.

$$s_{x,p} = \sqrt{\frac{b}{x} p(100 - p)} \qquad (9)$$

Here, x is the base of the percentage, p is the percentage ($0<p<100$), and b is parameter for the numerator of the percentage calculation. For the analysis using the SPD 2001 cross-sectional file data, the b parameters are provided in Table 3.

An illustration would be to suppose that, based on the final cross-sectional weight on the SPD 2001 cross-sectional file, an estimate of number of males aged 22 to 55 was 46,023,000. Among all the males in this age group, an estimate of 2.4 percent were unemployed. The b parameter associated with the numerator (the number of unemployed male) is 6,556 (from Table 3). Using Formula 9, the approximate standard error $s_{x,p}$ is

$$s_{x,p} = \sqrt{\frac{6{,}556}{46{,}023{,}000}(2.4)(100 - 2.4)} = 0.183\%$$

Consequently, the 90-percent confidence interval for the unemployment estimate is 2.10% to 2.70%.

To calculate the percentages of money, the formula is more complicated. A percentage of money will usually be estimated in one of two ways. It may be the ratio, p_M, of two aggregates as defined in Formula 10 below.

$$p_M = 100\left(\frac{X_A}{X_N}\right) \qquad (10)$$

or it may be the ratio, p_M, of two means with an adjustment, $\hat{p}_A$, for different bases as defined in Formula 11 below.

$$p_M = 100\left(\frac{\overline{X}_A}{\overline{X}_N}\right)\hat{p}_A \qquad (11)$$

where X_A and X_N in Formula 10 are aggregate money figures, $\overline{X}_A$ and $\overline{X}_N$ in Formula 11 are mean money figures, and $\hat{p}_A$ is the estimated number in Group A divided by the estimated number in Group N (i.e., the ratio of Group A to Group N). In either way of estimating p_M

(Formula 10 or 11), we estimate the standard error s_{p_M} of p_M using Formula 12 provided below (since p_M is expressed in percent, s_{p_M} is consistently expressed in percent as well).

$$s_{p_M} = 100 \times \sqrt{\left(\frac{\hat{p}_A \overline{X}_A}{\overline{X}_N}\right)^2 \left[\left(\frac{s_{\hat{p}_A} \div 100}{\hat{p}_A}\right)^2 + \left(\frac{s_{\overline{X}_A}}{\overline{X}_A}\right)^2 + \left(\frac{s_{\overline{X}_N}}{\overline{X}_N}\right)^2\right]} \qquad (12)$$

where $s_{\hat{p}_A}$ is the standard error of $\hat{p}_A$, $s_{\overline{X}_A}$ is the standard error of $\overline{X}_A$ and $s_{\overline{X}_N}$ is the standard error of $\overline{X}_N$. To calculate $s_{\hat{p}_A}$, use Formula 9 (which produces $s_{\hat{p}_A}$ in percent instead of in ratio like $\hat{p}_A$). The standard errors $s_{\overline{X}_A}$ and $s_{\overline{X}_N}$ are calculated using Formula 2.

Note that there is frequently some correlation among the characteristics estimated by $\hat{p}_A$, $\overline{X}_A$, and $\overline{X}_N$. These correlations, if present, will cause a tendency toward overestimates or underestimates, depending on the relative sizes of the correlations and whether they are positive or negative, respectively.

An illustration would be to suppose that, based on the data and the final cross-sectional weight on the SPD 2001 cross-sectional file, an estimated 8.8% of males aged 17 and over were Black in 2001, the mean annual earning of these Black males was $15,456 in 2000, the mean annual earning of all males aged 17 and over in 2001 was $22,932 in 2000, and the corresponding standard errors are 0.37 percent, $432, and $324, respectively. Then, the percent (p_M) of male earnings made by Blacks in 2000 per Formula 11 is

$$p_M = 100\left(\frac{15{,}456}{22{,}932}\right)(0.088) = 5.9\%$$

Using Formula 12, the approximate standard error, s_{p_M} is

$$s_{p_M} = 100 \times \sqrt{\left(\frac{(0.088)(15{,}456)}{22{,}932}\right)^2 \left[\left(\frac{0.0037}{0.088}\right)^2 + \left(\frac{432}{15{,}456}\right)^2 + \left(\frac{324}{22{,}932}\right)^2\right]} = 0.31\%$$

Standard Error of a Difference

The standard error s_{x-y} of a difference between two sample estimates *x* and *y* is equal to

$$s_{x-y} = \sqrt{s_x^{\,2} + s_y^{\,2} - 2rs_x s_y} \qquad (13)$$

where s_x and s_y are the standard errors of the estimates *x* and *y*. The estimates can be numbers, averages, percents, ratios, etc. The correlation between *x* and *y* is represented by *r* ($-1 \le r \le 1$). If *r* is assumed to be zero and the true correlation is really positive (negative), then this assumption will result in a tendency toward overestimates (underestimates) of the true standard error.
An illustration would be to suppose that we are interested in the difference in the average annual number of adult males (aged 17 and above in 2001) versus the corresponding adult females with annual cash income above \$17,000 in 2000. An estimate of the number of adult people in this income bracket has been obtained for both males and females. For females, the estimate is 1,619,000. A similar estimate for males is 2,198,000. The difference in estimates is 579,000.

The standard error of the adult female estimate is computed next. The *a* and *b* parameters from Table 3 for females are -0.0000648 and 6,556, respectively. Based on Formula 2, the standard error, s_x of the female estimate is

$$s_x = \sqrt{(-0.0000648)(1{,}619{,}000)^2 + (6{,}556)(1{,}619{,}000)} = 102{,}197$$

Similarly, the a and b parameters from Table 3 for males are -0.0000700 and 6,556, respectively. Based on Formula 2, the standard error, s_y of the male estimate is

$$s_y = \sqrt{(-0.0000700)(2{,}198{,}000)^2 + (6{,}556)(2{,}198{,}000)} = 118{,}625$$

Now, the standard error of the difference is computed using the above two standard errors. The correlation *r* for this example is assumed to be zero. The standard error, $s_{x\text{-}y}$ of the difference is computed by Formula 13 as shown below.

$$s_{x-y} = \sqrt{(102{,}197)^2 + (118{,}625)^2} = 156{,}577$$

Suppose that it is desired to test at the 10 percent significance level whether the number of 2001 adult males and females with 2000 annual cash income above \$15,000 were different, one can compare the difference of 579,000 to the product $1.645 \times 156{,}577 = 257{,}569$. Since the difference is larger than 1.645 times the standard error ($s_{x\text{-}y}$) of the difference, the data allow us to conclude that the number of adult males with annual cash income above $17,000 is significantly higher than the number of the adult females at the 10 percent confidence level.

Standard Error of a Median

The median quantity, X_{med}, of some item (characteristic), X, such as income for a given group of people, families, or households is that quantity such that at least half the group has as much or more and at least half the group has as much or less. The sampling variability of an estimated median $\hat{X}_{med}$ depends upon the form of the distribution of the item as well as the size of the group. To estimate the median (X_{med}) and the standard error of the median ($s_{X_{med}}$) the procedure described below may be used.

The median (X_{med}), like the mean, can be estimated using either data which has been grouped into intervals (e.g., income intervals) or ungrouped data. If grouped data are used, the median (X_{med}) is estimated using either Formula 15 or 16 with $p = 0.5$. If ungrouped data are used, the data records are ordered based on the value of the item (e.g., income level), then the estimated median is the value of the item such that the weighted estimate of 50 percent of the sub-population falls at or below that value and 50 percent is at or above that value. The method of standard error computation presented here requires the use of grouped data, because it is deemed easier to compute the median by grouping the data and then using Formula 15 or 16.

An approximate method for measuring the reliability of an estimated median ($\hat{X}_{med}$) is to determine a confidence interval about it. (See the section on "Confidence Intervals.") The following procedure (four steps) may be used to estimate the 68-percent confidence limits (i.e., approximately ± one standard error from the median) and hence the standard error (of a median) based on sample data.

<u>Step 1</u> - Determine, using Formula 9, the standard error ($s_{x,p=50}$) of an estimate of 50 percent of the group (sub-population).

<u>Step 2</u> - Subtract from and add to 50 percent the standard error determined in Step 1 to obtain the percentages associated with the lower and upper limits of the 68-percent confidence interval of the item. Namely, the smaller percentage is $50 - s_{x,p=50}$ percent, and the larger percentage is $50 + s_{x,p=50}$ percent.

<u>Step 3</u> - Using the distribution of the item within the group, calculate the quantity, X_{UCL}, of the item such that the percent of the group owning more of the item is equal to the smaller percentage $(50 - s_{x,p=50})$ found in Step 2. This quantity (X_{UCL}) will be the upper limit for the 68-percent confidence interval (assuming that the interval with higher item value is ranked at lower percentile as illustrated in Table 2.) In a similar fashion, calculate the quantity, X_{LCL}, of the item such that the percent of the group owning more of the item is equal to the larger percentage $(50 + s_{x,p=50})$ found in Step 2. This quantity (X_{LCL}) will be the lower limit for the 68-percent confidence

interval. (Note that a median computed from ungrouped data may or may not fall in this confidence interval).

Step 4 - Divide the difference between the two quantities (X_{UCL} and X_{LCL}) determined in Step 3 by two to obtain the standard error estimate ($s_{\hat{X}_{med}}$) of the median estimate ($\hat{X}_{med}$). Namely,

$$\hat{s}_{\hat{X}_{med}} = \frac{X_{UCL} - X_{LCL}}{2} \qquad (14)$$

To perform Step 3, it will be necessary to interpolate, which may be done using different methods. The most common is simple linear interpolation (Formula 15) and Pareto interpolation (Formula 16). The appropriateness of the method depends on the form of the distribution around the median. We recommend Pareto interpolation in most instances. Interpolation is used as follows. The quantity of the item, X_{pN} such that p percent own more of the item is

$$X_{pN} = A_1 \exp\left[\frac{\ln\left(\frac{pN}{N_1}\right)}{\ln\left(\frac{N_2}{N_1}\right)} \ln\left(\frac{A_2}{A_1}\right)\right] \qquad (15)$$

if Pareto Interpolation is indicated and

$$X_{pN} = \left[\frac{pN - N_1}{N_2 - N_1}\left(A_2 - A_1\right) + A_1\right] \qquad (16)$$

if linear interpolation is indicated, where N is the size of the group; A_1 and A_2 are the lower and upper bounds, respectively, of the interval in which X_{pN} falls; N_1 and N_2 are the estimated numbers of group members owning more than A_1 and A_2, respectively; *exp* refers to the exponential function; and *ln* refers to the natural logarithm function. One should note that a mathematically equivalent result is obtained by using common logarithms (base 10) and antilogarithms.

An illustration would be in order to calculate the standard error of a median, we return to the first example used to illustrate the standard error of a mean. As indicated in Table 2, the size (N) of the group is 39,923,000 and the median annual income estimate ($\hat{X}_{med}$) for the group falls in between

\$17,500 and \$19,999. With $p = 0.5$, $A_1 = \$1$, $s_{\hat{X}_{med}} = 7,500$, $A_2 = \$19,999$; $N_1 = 5,799,000 + 4,730,000 + ... +1,493,000 = 22,178,000$, and $N_2 = 4,730,000 + 3,723,000 + ... + 1,493 = 16,379,000$; the median annual income estimate, $\hat{X}_{med}$ for this group is computed using Formula 15 (with $p = 0.5$) to be \$18,331. The standard error estimate ($s_{\hat{X}_{med}}$) of the median annual income estimate is calculated using the above four step procedure as follows.

Step 1 - Using Formula 9 and the appropriate b parameter of 6,556, the standard error estimate of 50 percent on a base of 39,923,000 is about 0.64 percentage points, (i.e., $s_{x,p=50} = 0.64\%$).

Step 2 - Obtain the two percentages associated with the lower and upper limits of the 68-percent confidence: the smaller percentage = 50 - $s_{x,p=50}$ = 49.36 and the larger percentage = 50 + $s_{x,p=50}$ =50.64.

Step 3 - By examining Table 2, we see that the percentage 49.36 falls in the income interval from \$17,500 to \$19,999. Thus as determined previously, $A_1 = \$17,500$, $A_2 = \$19,999$, $N_1 = 22,178,000$, $N_2 = 16,379,000$, and $N = 39,923,000$ and $p = 49.36$. Based on Formula 15, the upper bound (X_{UCL}) of a 68-percent confidence interval for the median estimate ($\hat{X}_{med}$) is

$$X_{UCL} = 17,500 \exp\left[\frac{\ln\left(\dfrac{0.4936 \times 39,923,000}{22,178,000}\right)}{\ln\left(\dfrac{16,379,000}{22,178,000}\right)} \ln\left(\frac{19,999}{17,500}\right)\right] = \$18,435$$

Also by examining Table 2, the 50.64 percent fall in the same income interval. Thus, A_1, A_2, N_1, and N_2 are the same as above, but $p = 0.5064$. The lower bound (X_{LCL}) of a 68-percent confidence interval for the median ($\hat{X}_{med}$) is

$$X_{LCL} = 17,500 \exp\left[\frac{\ln\left(\dfrac{0.5064 \times 39,923,000}{22,178,000}\right)}{\ln\left(\dfrac{16,379,000}{22,178,000}\right)} \ln\left(\frac{19,999}{17,500}\right)\right] = \$18,228$$

<u>Step 4</u> - Based on Formula 14, the standard error estimate ($s_{\hat{X}_{med}}$) of the median annual income estimate ($\hat{X}_{med}$) is

$$s_{\hat{X}_{med}} = \frac{\$18{,}435 - \$18{,}228}{2} = \$104$$

If the linear interpolation is used, the median is then estimated using Formula 16 to be $18,455 and the 68-percent confidence interval of the estimated median is from $18,345 to $18,565. The standard error estimate is $110.

Standard Error of Ratio of Means or Medians

The standard error for a ratio of means or medians is approximated by Formula 17 provided below.

$$s_{\frac{X}{Y}} = \sqrt{\left(\frac{X}{Y}\right)^2 \left[\left(\frac{s_X}{X}\right)^2 + \left(\frac{s_Y}{Y}\right)^2\right]} \tag{17}$$

where X and Y are the means or medians, and s_X and s_Y are their associated standard errors. Formula 17 assumes that the means or medians are not correlated. If the correlation between the population means or medians estimated by X and Y are actually positive (negative), then this procedure will tend to produce overestimates (underestimates) of the true standard error for the ratio of means or medians.

Table 1 - Reference months and interview months of the SIPP 1992 Panel, SIPP 1993 Panel, SPD Bridge (1997), and SPD 1998 to SPD 2002 survey samples.

Survey Sample	Months of Interview	Reference Months
SIPP Panel 1992	February 1992 - April 1995	October 1991 - March 1995
SIPP Panel 1993	February 1993 - January 1996	October 1992 - December 1995
SPD Bridge (1997)	April 1997 - June 1997	January 1996 - December 1996: rolled-up as 1996 yearly data
SPD 1998	May 1998 - July 1998	January 1997 - December 1997: rolled up as 1997 yearly data
SPD 1999	May 1999 - July 1999	January 1998 - December 1998: rolled up as 1998 yearly data
SPD 2000	May 2000 - July 2000	January 1999 - December 1999: rolled up as 1999 yearly data
SPD 2001	May 2001 - July 2001	January 2000 - December 2000: rolled up as 2000 yearly data
SPD 2002	May 2002 - June 2002	January 2001 - December 2001: rolled up as 2001 yearly data

Table 2 - Illustrative example: distribution of annual income among people 25 to 34 years old.

	Total Number of People	Number of People in Annual Income Interval												
		Under $5000	$5000 to $7499	$7500 to $9999	$10000 to $12499	$12500 to $14999	$15000 to $17499	$17500 to $19999	$20000 to $29999	$30000 to $39999	$40000 to $49999	$50000 to $59999	$60000 to $69999	$70000 and Over
Number of People (in Thousands)	39923	1371	1651	2259	2734	3452	6278	5799	4730	3723	2591	2619	1223	1493
Percent with at Least as Much as Lower Bound of Interval	N/A	100.0	96.6	92.4	86.8	79.9	71.3	55.6	41.0	29.2	19.9	13.4	6.8	3.7

Note: This table contains a fictitious distribution of annual income and is used only to illustrate standard error calculations.

Table 3 - SPD generalized variance parameters for estimates of various characteristics using the final cross-sectional weight (SPDCSW01) on the SPD 2001 cross-sectional file and the corresponding design effects (Deff).

Characteristics	Generalized Variance Parameters		Design Effects (Deff)
	a	b	
TOTAL OR WHITE PEOPLE			
16+ Program Participation and Benefits, Poverty (3)*			
Both Sexes	-0.0000533	9,185	3.92
Male	-0.0001099	9,185	3.92
Female	-0.0001034	9,185	3.92
16+ Income and Labor Force (5)*			
Both Sexes	-0.0000337	6,556	1.34
Male	-0.0000700	6,556	1.34
Female	-0.0000648	6,556	1.34
16+ Pension Plan** (4)*			
Both Sexes	-0.0000615	12,007	2.45
Male	-0.0001280	12,007	2.45
Female	-0.0001186	12,007	2.45
All Others*** (6)*			
Children Aged Less Than 18			
Both Sexes	-0.0000448	11,502	4.86
Male	-0.0000917	11,502	4.86
Female	-0.0000875	11,502	4.86
Adults Aged 18 and Over			
Both Sexes	-0.0000929	23,846	4.86
Male	-0.0001902	23,846	4.86

Characteristics	Generalized Variance Parameters		Design Effects (Deff)
	a	b	
Female	-0.0001815	23,846	4.86
BLACK PEOPLE			
Poverty (1)*			
Both Sexes	-0.0002410	7,835	3.34
Male	-0.0005159	7,835	3.34
Female	-0.0004524	7,835	3.34
All Others*** (2)*			
Children Aged Less Than 18			
Both Sexes	-0.0001309	4,255	1.80
Male	-0.0002802	4,255	1.80
Female	-0.0002457	4,255	1.80
Adults Aged 18 and Over			
Both Sexes	-0.0002714	8,821	1.80
Male	-0.0005809	8,821	1.80
Female	-0.0005094	8,821	1.80
HOUSEHOLDS			
Total or Whites	-0.0000703	6,721	1.65
Black	-0.0004191	4,644	1.14

* For cross-tabulations, use the *a* and *b* parameters of the characteristic with the smaller number within the parentheses.

** Use the "16+ Pension Plan" parameters for pension plan tabulations of people aged 16+ in the labor force. Use the "All Others" parameters for retirement tabulations, 0+ program participation, 0+ benefits, 0+ income, and 0+ labor force tabulations, in addition to any other types of tabulations not specifically covered by another characteristic in this table.

*** Use the "All Others" parameters for any type of tabulation not specifically covered by another characteristic in this table.